NFL Adjusted Plus-Minus Special Teams Models

Metrics Assessing Player's EPA and Penalty Contribution During Punts

Consultant: Matt Johnson
Client: NFL Big Data Bowl
Advisors: Dr. Steve Marron, Ph.D. & Dr. Perry Haaland, Ph.D.

16 March, 2022

Abstract

Introduction

This paper introduces two new metrics for evaluating player performance during punting situations in the NFL: Regularized Adjusted Expected Points Added Contribution (RAEPAC) and Regularized Adjusted Penalty Yard Contribution (RAPYC). To build these new metrics, I adapt a statistic historically used in the NBA called Adjusted Plus-Minus (APM). This metric adjusts the conventional Plus-Minus metric to control for every player on the court during a stint where the same 10 players are on the court. In the NFL punt return case, this is simply the 22 players on the field during a punt. In classic APM for basketball, the variable of interest is the points scored during the stint, however this will not work for football since points are rarely scored during punts. In RAEPAC, I use Expected Points Added (EPA) as the variable of interest while in RAPYC I use penalty yards. EPA calculation is further explained in the Methodology section. Specifically, to calculate RAEPAC I use Ridge regression to to get regression coefficients for each player that was on the field during a punt. To calculate RAPYC, I use a Zero-Inflate Poisson regression to get coefficients for each player. These regression coefficients are what define these two new metrics. The model choices will be discussed further in the Methodology section. These metrics can be used to rank players regardless of their position on how much they positively contribute towards the success of their team on punting situations. These types of metrics could be extremely useful for on-field decision makers when deciding who should be on the field during crucial punting plays. Also, these metrics can be aggregated on a team level to rank how successful teams are during punts. Assessing players regardless of their position could change the way coaches think about their personnel and shed light on players that may not get enough credit for their contribution.

Data

The final dataframe used for analysis has 5890 plays with 2005 unique players. Each row of is a play from 2018-2020 where the team lined up to punt the ball away and each column is a player that could have been on the field. Each cell of the dataframe represents if the player was on the field during that play or not. A player on the punting team is given a -1 and each player on the receiving team is given a 1. Players that were not on the field during that play are coded as a 0. This follows the methodology first employed by Dan Rosenbaum for computing APM is basketball.

This dataframe is inherently rank deficient and therefore I cannot use standard linear regression, rather I use a penalized approach so the sample covariance matrix was invertible. The dataframe is rank deficient because if I remove a players column, I still know whether he was on the field on offense or defense by simply

using the rest of the columns. According to this, the matrix does not have full rank. Below is a glimpse of the dataframe.

PlayId	EPA	Pen.Yrds	32361	38707	40008	40191	40408	42450	42515
2018090600_366	0.75865422	-15	-1	1	1	1	-1	1	-1
2018090600_872	-0.41094785	0	0	-1	0	-1	1	0	1
2018090600_973	-0.51340565	0	-1	1	0	1	-1	1	-1
2018090600_1850	-0.19499974	0	0	-1	0	-1	1	0	1
2018090600_1989	-0.34500556	12	0	-1	0	-1	1	0	1
2018090600_2406	0.08443995	0	0	-1	0	-1	1	0	1

Methodology

Results

Conclusion & Future Research

References